

What is claimed is:

1. A method of detecting a mismatch in any of a plurality of DNA duplexes of distinct nucleic acid sequence, said duplexes formed in a single hybridization reaction, comprising:
detecting, for any of said duplexes, an alteration in a characteristic of a cell, said alteration effected by the *in vivo* mismatch corepair of a marker that is present together with said duplex in a vector within said cell, said corepair being initiated by a mismatch in said duplex.
2. The method of claim 1, wherein said cell is a yeast cell.
3. The method of claim 1, wherein said cell is a bacterial cell.
4. The method of claim 3, wherein said mismatch corepair is mediated by a bacterial methyl mismatch repair system.
5. The method of claim 4, wherein said bacteria are *E. coli*, said methyl mismatch repair system is *E. coli* *dam*-directed mismatch repair, and said vector is hemimethylated.
6. The method of claim 5, wherein said marker is a heteroduplex of a first and a second nucleic acid strand, said strands differing in sequence by at least five consecutive nucleotides.
7. The method of claim 6, wherein said initiating mismatch is directly adjacent to no more than 3 additional mismatches
8. The method of claim 1, wherein said plurality includes at least least 10 duplexes of distinct nucleic acid sequence.

9. The method of claim 8, wherein said plurality includes at least least 1,000 duplexes of distinct nucleic acid sequence.
10. The method of claim 9, wherein said plurality includes at least least 10,000 duplexes of distinct nucleic acid sequence.
11. The method of claim 10, wherein said plurality includes at least least 100,000 duplexes of distinct nucleic acid sequence.
12. The method of claim 1, wherein said plurality includes nucleic acid sequences derived from a eukaryote.
13. The method of claim 12, wherein said eukaryote is a mammal.
14. The method of claim 13, wherein said mammal is a human.
15. The method of claim 14, wherein said plurality includes nucleic acid sequences derived from the coding region of a human gene.
16. The method of claim 15, wherein said plurality includes nucleic acid sequences derived from noncoding regions of the human genome.
17. The method of claim 1, further comprising the antecedent step of forming said plurality of DNA duplexes by annealing first nucleic acid strands, said first strands including at least one nucleic acid sequence, to second nucleic acid strands, said second strands including a plurality of distinct nucleic acid sequences.
18. The method of claim 17, wherein said plurality of second nucleic acid strands are derived from a common source.

19. The method of claim 18, wherein said common source is genomic DNA or cDNA from a single individual.

20. The method of claim 17, wherein said plurality of second nucleic acid strands are derived from a pooled source.

21. The method of claim 20, wherein said source is pooled from family members.

22. The method of claim 1, further comprising a final step of isolating said duplexes from said cells.

23. The method of claim 21, wherein said cells are yeast cells.

24. The method of claim 21, wherein said cells are bacterial cells.

25. The method of claim 24, wherein said bacterial cells are the MUTATION SORTER™ (MS) strain.

26. A kit for detecting a mismatch in any of a plurality of DNA duplexes of distinct nucleic acid sequence, comprising:

a first vector and a second vector, each such vector including an origin suitable for replication in a cell and a sequence that encodes a marker, the said first and second vector marker encoding sequences differing from one another, said sequence difference capable of undergoing but not initiating *in vivo* mismatch repair.

27. A method of identifiably detecting a mismatch in any of a plurality of DNA duplexes of distinct nucleic acid sequence, comprising:

phenotypically sorting from said plurality of distinct duplexes those capable of initiating a mismatch corepair event *in vivo*; and then

identifying the duplexes present in said phenotypically sorted population,

wherein identification is effected by identifying at least one genotypically detectable genetic element uniquely linked to each said phenotypically sorted duplex.

28. The method of claim 27, wherein each of said at least one genotypically detectable genetic elements is a nucleic acid sequence tag, each of said sequence tags being unique among said plurality of sequence tags, and wherein said sorted duplexes are identified by specific hybridization of said sequence tags, tagged duplexes, or nucleic acids derived therefrom, to a microarray having probes complementary to said plurality of sequence tags.

29. The method of claim 28, wherein each of said plurality of distinct sequence-tagged duplexes is itself further linked to at least one priming sequence, said at least one priming sequence being sufficient to allow enzymatic amplification of the tagged duplex linked thereto.

30. The method of claim 28 or claim 29, wherein each of said sequence tags is at least 17 nucleotides in length.

31. The method of claim 30, wherein each of said sequence tags is at least 20 nucleotides in length.

32. The method of claim 31, wherein each of said sequence tags is at least 25 nucleotides in length.

33. The method of claim 27, wherein at least one among said plurality of distinct DNA duplexes has at least one strand identical in sequence to a naturally-occurring genomic sequence.

34. The method of claim 33, wherein each of said at least one duplexes having sequence identical to naturally-occurring genomic sequence is obtained by

amplification from a genomic template.

35. The method of claim 33 or claim 34, wherein said genomic sequence is a eukaryotic genomic sequence.

36. The method of claim 35, wherein said eukaryotic genomic sequence is selected from the group consisting of: yeast genomic sequence, plant genomic sequence, and mammalian genomic sequence.

37. The method of claim 36, wherein said eukaryotic genomic sequence is mammalian genomic sequence.

38. The method of claim 37, wherein said mammalian genomic sequence is selected from the group consisting of: murine genomic sequence, rattus genomic sequence, and human genomic sequence.

39. The method of claim 38, wherein said mammalian genomic sequence is human genomic sequence.

40. The method of claim 33, wherein said plurality of distinct DNA duplexes includes at least two duplexes having the sequence of different allelic variants of a single genomic locus.

41. The method of claim 40, wherein said plurality of distinct DNA duplexes includes at least three duplexes having the sequence of different allelic variants of a single genomic locus.

42. The method of any one of claims 27, 33 or 40, wherein said plurality of distinct DNA duplexes includes at least 10 distinct DNA duplexes.

43. The method of claim 32, wherein said plurality of distinct DNA duplexes

includes at least 100 distinct DNA duplexes.

44. The method of claim 43, wherein said plurality of distinct DNA duplexes includes at least 1000 distinct DNA duplexes.

45. The method of claim 44, wherein said plurality of distinct DNA duplexes includes at least 5000 distinct DNA duplexes.

46. The method of claim 45, wherein said plurality of distinct DNA duplexes includes at least 10,000 distinct DNA duplexes.

47. The method of claim 27, further comprising:
phenotypically sorting from said plurality of distinct duplexes those that are incapable of initiating a mismatch corepair event *in vivo*; and then
identifying duplexes present in said further phenotypically sorted population, wherein said identifying is identification of at least one genotypically detectable genetic element uniquely linked to each said duplex.

48. An improved standard vector for use in mismatch repair detection, the improvement comprising:
uniquely linking a genotypically detectable genetic element to the vector's standard sequence.

49. An improved standard vector for use in mismatch repair detection, the improvement comprising:
operatively linking the phenotypically sortable genetic element to a regulated strong promoter and a heterologous ribosomal binding site.

50. The vector of claim 48 or claim 49, wherein said phenotypically sortable genetic element encodes a recombinase.

51. The vector of claim 49, wherein said regulated strong promoter is a T7 promoter.

52. The vector of claim 51, further comprising a genotypically detectable genetic element uniquely linked to said standard sequence.

53. An improved mismatch repair vector, the improvement comprising:
positioning the mismatch in the phenotypically sortable genetic element to be no more than 200 nucleotides from the test duplex.

54. A method of preparing standard vectors for mismatch repair detection, comprising:

propagating a double-stranded closed circular vector in a bacterial strain under conditions permissive for *dam* expression, said vector comprising a plasmid origin of replication, a filamentous phage origin of replication, a standard sequence, and a phenotypically sortable genetic element;

changing said propagation conditions to be nonpermissive for *dam* expression; and then

rescuing closed circular single stranded nucleic acids from said propagated vector by infection of said bacterial strain with helper phage.

55. A method of identifying alleles of a genomic locus, comprising:
duplexing a single-stranded standard for each said locus with nucleic acids pooled from a plurality of separate individuals;
phenotypically sorting from said plurality of distinct duplexes those capable of initiating a mismatch corepair event *in vivo*; and then
identifying the duplexes present in said phenotypically sorted population,
said identifying being identification of at least one genotypically detectable genetic element uniquely linked to each said duplex.